

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Pertumbuhan data ternyata sangat mempengaruhi perkembangan volume dan jenis data yang terus meningkat di dunia maya. Jenis data, mulai dari data yang berupa teks, gambar atau foto, video hingga bentuk data-data lainnya membanjiri sistem komputasi. Tidak hanya jumlah data yang di-generate yang bertambah, tetapi tingkat kenaikan juga bertambah cepat [1].

Pertumbuhan data ini oleh banyak perusahaan diistilahkan dengan *big data*. Belum ada definisi pasti mengenai *big data*. Namun, oleh banyak perusahaan *big data* diartikan dengan tantangan yang terus tumbuh yang harus dihadapi dan diselesaikan oleh organisasi ataupun perusahaan yang berupa sumber data dalam jumlah besar dan pertumbuhan data yang cepat, atau informasi yang berisi analisis dengan lingkup data yang tidak sederhana.

Big data memiliki beberapa dimensi atau karakteristik yang dideskripsikan ke dalam *volume*, *variety*, dan *velocity*. Dimensi ini dapat mengklasifikasikan masalah yang ditimbulkan oleh *big data*. Internet memfasilitasi pertumbuhan data yang besar, karenanya beberapa jejaring sosial seperti twitter dapat menerima kicauan dari penggunaanya yang berjumlah lebih dari 400 juta dalam sehari. Permasalahan *volume* data seperti ini tidak dapat lagi diselesaikan dengan metode penghitungan yang tradisional, oleh karena itu dibutuhkan metode baru yang dapat mengatasi masalah *volume* data ini. Salah satu contoh aplikasi yang menggunakan metode *big data* adalah apache Hadoop. Aplikasi ini bersifat *free* dan *open source*. Banyak pengembangan aplikasi *big data* lain yang pada dasarnya menggunakan aplikasi Hadoop, contohnya IBM.

Hadoop bisa dijalankan disatu computer saja (*single-node*) ataupun dalam cluster yang berisi lebih dari satu computer (*multi-node*) dengan menggunakan *platform* java. *Single-node* biasanya digunakan hanya untuk latihan atau percobaan saja. *Multi-node* merupakan gabungan dari *server* dan *client*. Dengan menggunakan metode *multi-node* tersebut, Hadoop menyatakan dapat memproses data dengan lebih baik.

Mengetahui perkembangan data yang telah terjadi dan kebutuhan untuk aplikasi yang dapat mengelola data tersebut, tentu membutuhkan pemahaman mengenai Hadoop yang merupakan salah satu perangkat lunak yang dapat menangani permasalahan tersebut. Tetapi, faktanya masih sangat sulit untuk menemukan informasi mengenai Hadoop. Maka dari itu akan diadakan pembuktian terutama pada *client* bagaimana data akan dikelola oleh Hadoop.

1.2 Rumusan Masalah

Berikut masalah – masalah yang mungkin akan terjadi dan perlu dijawab. Masalah – masalah yang ada sebagai berikut:

1. Bagaimana karakteristik penerimaan data dari *server* ke *client*?
2. Apakah ukuran data mempengaruhi waktu penerimaan data?
3. Apakah ukuran *block* data mempengaruhi waktu penerimaan data?

1.3 Tujuan Pembahasan

Tujuan dari penelitian ini adalah sebagai berikut:

1. Mengamati bagaimana karakteristik penerimaan data dari *server* ke *client*.
2. Mengamati dan menganalisis pengaruh ukuran data terhadap waktu penerimaan data.
3. Mengamati dan menganalisis pengaruh ukuran *block* data terhadap waktu penerimaan data.

1.4 Ruang Lingkup

1. Penelitian dibatasi pada sisi *client*, dengan cara mengamati karakteristik pada *client* pada saat distribusi data.
2. Data dibatasi dengan data yang berukuran 800MB, 1.2GB, dan 2 GB.
3. Percobaan dilakukan dengan menggunakan metode *multi-node*.
4. Aplikasi Hadoop akan digunakan pada mesin virtual Oracle VirtualBox versi 5.0.0 dengan spesifikasi CPU single core, RAM 786 MB, sistem operasi Ubuntu 14.04 LTE.

5. Komputer yang digunakan untuk menjalankan mesin virtual memiliki spesifikasi, prosesor Intel Core i5-2320 3.00GHz, RAM 4 GB, dan sistem operasi Windows 7.
6. Ukuran *block* data yang digunakan dibatasi dengan ukuran 64 MB, 128 MB, dan 256 MB.

1.5 Sumber Data

Penelitian ini dilakukan dengan menggunakan data sekunder. Metode yang digunakan dalam membantu penulisan tugas akhir ini adalah:

1. Penelitian dalam bentuk pengujian *platform* perangkat lunak Hadoop melalui tahapan penginstalan, konfigurasi dan pengujian. Dalam pengembangan pengujian Hadoop ini, menggunakan studi literatur yang berkaitan dengan Hadoop dan *big data*.
2. Melakukan percobaan dari skenario yang telah dibuat.
3. Melakukan pengamatan, pengambilan data, analisis dan pengambilan kesimpulan dari hasil pengujian skenario yang telah dilakukan.

1.6 Sistematika Penyajian

Dalam proses penyusunan laporan, sistematika penulisan yang akan digunakan adalah sebagai berikut:

BAB I PENDAHULUAN

Bab ini berisi tentang latar belakang, rumusan masalah, tujuan pembahasan, ruang lingkup penelitian, sumber data, dan sistematika penyajian dari proyek tugas akhir.

BAB II DASAR TEORI

Bab ini berisi penjelasan teori tentang big data, HDFS, dan MapReduce.

BAB III ANALISA DAN RANCANGAN SISTEM

Topologi menggunakan satu buah komputer server dan dua buah komputer *client*. Setiap skenario menggunakan *file* berukuran 800 MB, 1.2 GB, dan 2 GB.

BAB IV IMPLEMENTASI

Bab ini berisi hasil implementasi dari awal penginstalan dan dokumentasi setiap kali melakukan konfigurasi perangkat lunak.

BAB V PENGUJIAN

Bab ini berisi hasil percobaan dari setiap skenario yang telah dibuat.

BAB VI SIMPULAN DAN SARAN

Bab ini berisi simpulan dari hasil penelitian tugas akhir dan saran yang dapat digunakan pada penelitian selanjutnya.

